

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-301650

(43)Date of publication of application : 28.10.1994

(51)Int.Cl.

G06F 15/16

G06F 13/00

H04L 12/56

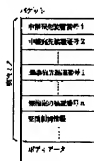
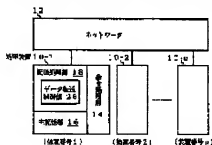
(21)Application number : 05-051186

(71)Applicant : NATL AEROSPACE LAB  
FUJITSU LTD

(22)Date of filing : 12.03.1993

(72)Inventor : MIYOSHI HAJIME  
FUKUDA MASAHIRO  
IKEDA MASAYUKI  
TAKAMURA MORIYUKI

## (54) DATA TRANSFER REPEATING MECHANISM



## (57)Abstract:

**PURPOSE:** To simplify a communication procedure and to reduce the overhead of processing by efficiently executing repeating transfer through other processors in a data transfer repeating mechanism for mutually executing data communication through other processors among processors provided in a data processing system such as a parallel computer system.

**CONSTITUTION:** Each of plural processors 10-1 to 10-n is provided with a data transfer control part 26 for specifying the device numbers of one or more through processors in addition to the device number of the final destination processor in the header part of a transmission packet and returning a receiving packet to a network 12 by setting up the next through processor to the next destination without processing body data at the time of detecting that its own processor is not the final destination processor based upon device

numbers in the header data of the receiving packet or returning the packet to a network 12 by setting up the final destination processor to the next destination when there is not next through processor.

## \* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.\*\*\* shows the word which can not be translated.

3.In the drawings, any words are not translated.

---

CLAIMS

---

[Claim(s)]

[Claim 1]Two or more processing units (10-1 - 10-n) provided with an instruction-processing part part (14), a main memory part (16), and a transmission treating part (18) are connected to a network (12) provided with a switching circuit. A header unit in which a transmission treating part (18) of a transmission source processor includes an address processing unit, the amount of data transmitting, a memory access attribute in a transmission source processor, and information about specification of processing in an address processing unit at least, A packet which contains body data at least when quantity of body data specified by this header unit is not zero is sent to said network (12). A switching circuit of said network (12) connects a switching circuit according to directions of an address processing unit contained in said packet, and in a transmission treating part (18) of an address processing unit. With reference to control information on reception included in a header unit when quantity of body data which said transmission source processor sent is not zero, send body data to an instruction-processing part (14) and a main memory part (16), and communication between processing units is performed. In a data processing system which performs communication between processing units with reference to control information included in a header unit when quantity of body data is zero, to each of said processing unit (10) at the time of transmission to a specific processing unit. A packet which specified the device number of one or more course processing units in a header unit in addition to the device number of a final destination processing unit is transmitted to a network (12). At the time of packet reception from a network (12), from the device number in a header unit of this receive packet, when a self-device is not a final destination processing unit, A data transfer relay mechanism providing a transfer controlling part (26) which returns a packet for a final destination processing unit to a network (12) as a following address when the following course processing unit is made into the following address, without processing body data or there is no following course processing unit.

[Claim 2]In the data transfer relay mechanism according to claim 1, a transfer controlling part (22) of a transmission source processor, A data transfer relay mechanism making a packet reach a final destination processing unit while a packet which stored the device number of all the course processing units to a final destination processing unit in a header unit was transmitted to said network and the device number of all the course processing units had been included in a header unit.

[Claim 3]In the data transfer relay mechanism according to claim 1 or 2, said transfer controlling part (22)

has the data transfer processing demand from an instruction-processing part (14), And a data transfer relay mechanism giving priority to transmission of a packet which relays a self-processing unit when a packet which requires data transfer which goes via a self-processing unit from other processing units simultaneously has been sent.

[Claim 4]In the data transfer relay mechanism according to claim 1 or 2, information about a priority of data transfer is included in HETTA data of said packet, When a packet which requires data transfer which goes via a self-processing unit from other processing units simultaneously with a data transfer processing demand from an instruction-processing part (14) has been sent, said transfer controlling part (22), A data transfer relay mechanism which compares a priority of transmission and to which a priority of a relay packet is characterized by giving priority to transmission of a packet which is the same as a priority of a transmitting packet, or relays a self-processing unit when high.

[Claim 5]In the data transfer relay mechanism according to claim 1 to 4, said transfer controlling part (22), It has a translator which searches for a real unit number of a final destination processing unit, and a real unit number of a course processing unit from a logical unit number of a final destination processing unit specified by transmission cue on a main memory part (16), A data relay control device generating a packet which includes a real unit number of a final destination device and a course processing unit acquired by this translator in a header unit.

## DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Industrial Application]This invention relates to the data transfer relay mechanism which communicates data mutually via other devices between the processing units formed in data processing systems, such as a parallel computer system.

[0002]

[Description of the Prior Art]In the data communications between the processing units in the data processing system which connected two or more processing units in the network, when an obstacle occurs in the data transfer path between two processing units, relay transfer which transmits data via other processing units is performed. For example, when it sends out the contents of the memory location of the processing unit A to other processing units B and a direct data transfer path cannot use it from the processing unit A for the reasons of failure etc. to the processing unit B, The processing unit A will perform data transfer to the processing unit C, and will perform data transfer from the processing unit C to the processing unit B after that.

[0003]It is more effective to perform relay transfer which goes via other processing units, also when a direct data transfer path is dramatically congested from the processing unit A to the processing unit B and the course of the processing unit A to the processing unit B has become empty with the network characteristic. If it is in a device conventionally, about the relay transfer sent to the processing unit B via the processing unit C from the processing unit A. After the supervisor program of the processing unit A requires

reservation of the memory area for transmission of the main memory of the processing unit C of a relay destination, It stores in the memory area which transmitted and secured data to the processing unit C from the processing unit A, and after the supervisor program of the processing unit C next requires reservation of the memory area for transmission of the main memory of the processing unit B, data transfer to the processing unit B is performed.

[0004]

[Problem(s) to be Solved by the Invention]However, in such conventional relay transfer. The supervisor program of the processing unit A of a transmitting agency must secure a memory area to the main memory of the processing unit C of a relay destination, must transmit data, must wait for the end of the data transfer to the processing unit C from the processing unit A, and data transfer to the processing unit B must be performed from the processing unit C.

[0005]For this reason, the overhead of the processing which data transfer takes is large, and a memory resource is also required. In order to save a memory resource and to have to transmit in several steps, the procedure of data transfer becomes still more complicated and an overhead becomes very large. This invention was made in view of such a conventional problem, and relates to the data transfer relay mechanism which performs efficiently relay transfer which went via other devices, and aimed at simplification of a communication procedure, and reduction of the overhead of processing.

[0006]

[Means for Solving the Problem]Drawing 1 is a principle explanatory view of this invention. This invention is first aimed at data processing systems, such as a multiprocessor system which connected two or more processing units 10-1 provided with the instruction-processing part part 14, the main memory part 16, and the transmission treating part 18 - 10-n to the network 12 provided with a switching circuit.

[0007]A header unit in which the transmission treating part 18 of a transmission source processor includes an address processing unit, the amount of data transmitting, a memory access attribute in a transmission source processor, and information about specification of processing in an address processing unit at least, A packet which contains at least body data specified by a header unit is transmitted to the network 12. A switching circuit of the network 12 connects a switching circuit according to directions of an address processing unit contained in a transmitting packet.

[0008]In the transmission treating part 18 of an address processing unit, when quantity of body data of a packet which a transmission source processor sent is not zero, with reference to information on processing included in a header unit, body data is sent to the instruction-processing part 14 or the main memory part 16, and is processed. If it is in a data transfer relay mechanism of this invention for a data processing system which performs communication between such processing units, To each of the processing unit 10-1 - 10-n, at the time of transmission to arbitrary processing units. A packet which specified the device number of one or more course processing units in a header unit in addition to the device number of a final destination processing unit is transmitted to the network 12. At the time of packet reception from the network 12. From the device number in header data of a receive packet, when a self-device is not a final destination processing unit, When the following course processing unit was made into the following address, without processing body data or there was no following course processing unit, the control data transfer 26 which returns a packet for a final destination processing unit to the network 12 as a following address was

formed.

[0009]Here the control data transfer 26 of a transmission source processor, A packet which stored the device number of all the course processing units to a final destination processing unit in a header unit is transmitted to the network 12, and a packet is made to reach a final destination processing unit, while the device number of all the course processing units had been included in a header unit. The control data transfer 26 has the data transfer processing demand from the instruction-processing part 14, and when a packet which requires data transfer which goes via a self-processing unit from other processing units simultaneously has been sent, it gives priority to transmission of a packet which relays a self-processing unit.

[0010]Information about a priority of data transfer is included in header data of a packet, When a packet which requires data transfer which goes via a self-processing unit from other processing units simultaneously with a data transfer processing demand from the instruction-processing part 14 has been sent, the control data transfer 26, A priority of transmission is compared, and it is the same as a priority of a transmitting packet, or a priority of a relay packet gives priority to transmission of a packet which relays a self-processing unit, when high.

[0011]On the other hand, the control data transfer 26 is provided with a translator which searches for a real unit number of a final destination processing unit, and a real unit number of a course processing unit from a logical unit number of a final destination processing unit specified by the transmission cue 20 on the main memory part 16, It may be made to generate a packet which includes a real unit number of a final destination device and a course processing unit acquired by this translator in a header unit.

[0012]

[Function]According to the data transfer relay mechanism of this invention provided with such composition, the next operation is obtained. If it is in the control data transfer 26 of a transmission source processor first, the packet which put the device number of the processing unit via which it goes besides the device number of a final destination processing unit in the header unit (control information part) is generated, and it sends to the network 12.

[0013]The control data transfer 26 provided in the processing unit of the receiver will distinguish whether it is a packet which reads the destination designation number in a header unit, relays a self-processing unit, and makes other processing units an address, if the header unit of a receive packet goes into the receive data buffer 24 from the network 12. The usual reception is started when it is a packet which makes a self-processing unit a final destination. When it is a packet which relays a self-processing unit and makes other processing unit devices a final destination, after changing the device number of the present address processing unit in a header unit to the device number of the relay processing device received next or a final destination processing unit, it retransmits a message to a network.

[0014]In what is performing conversion for a real processing unit number from the logical unit number of the transmission cue on the main memory part 16 by the control data transfer 26. The translator which can define the real processing unit number and the real processing unit number of a final destination which are relayed to one logical unit number is provided, and it can process at high speed, without being conscious of the relay data transfer which had a complicated procedure by using this translator by a user program.

[0015]

[Example]Drawing 2 is an example lineblock diagram of the multiprocessor system with which the data transfer relay mechanism of this invention is applied, the processing unit 10-1 known as a processor element (PE) if it was in this example, and ... 10-n is connected to the network 12 and it is made to perform data transfer between the processing unit 10-1 - 10-n via the network 12.

[0016]That is, each of the processing unit 10-1 - 10-n creates the packet which comprised a header unit and body data, and transmits to the network 12. Here, in the header unit of a packet, the access attribute of the memory in the device number of an address processing unit, the amount of data transmitting, and a transmission source processor and the information about specification of the processing in an address processing unit are included at least.

[0017]The network 12 comprises a crossbar switch and a multi stage switching circuit network, and performs switch connection according to directions of the address processing unit contained in the packet which the processing unit 10-1 - 10-n sent. The processing unit which received the packet via the network 12 performs reception with reference to the control information on the reception in the address processing unit contained in the header unit of a packet.

[0018]If it is in this invention in addition to the function about the usual packet sending and receiving of such a processing unit 10-1 - 10-n, If it is in a transmission source processor, it adds to the device number of a final destination processing unit into the header unit of a transmitting packet, The following course processing unit is made into the following address, without putting in the device number of one or more course processing units, and processing body data from the device number in the header unit of a receive packet, when a self-processing unit is not a final destination processing unit, Or when there is no following course processing unit, the function to return a packet to a network as a following address is provided in a final destination processing unit.

[0019]Drawing 3 is an example lineblock diagram of the processing unit of drawing 2, and has the instruction-processing part 14, the main memory part 16, and the transmission treating part 18. The transmission cue 20 which stored the control information used for packet transfer in the data format of the packet header part is formed in the main memory part 16. The control data transfer 26 which realizes the function as a data relay transfer mechanism of this invention is formed in the transmission treating part 18.

[0020]Drawing 4 is an example lineblock diagram showing the details of the transmission treating part 18 of drawing 3. In drawing 4, the transmission treating part 18 comprises the data buffer 24, the control data transfer 26, the transmission counter 28, the transmission cue base address register 30, the transmission cue read-out pointer 32, the transmission cue writing pointer 34, and the main storage access control section 36.

[0021]Drawing 5 shows the contents of the transmission cue 20 created on the main memory part 16 shown in drawing 3. In drawing 5, if the information about packet transfer control is stored in the transmission cue on main memory in the data format of a header unit at a packet unit and the header unit 40 for one packet of a graphic display is taken for an example, body data length, a transmission address, a recipient address, etc. will be formed following the destination information field 42. The processing unit 10-1 shown in drawing 2 - the device number 1 of 10-n - n are registered into the destination information field 42 as reception device specification 1-n. Each of the reception device specification 1 - n comprises the BARIDDO bit v and the processing unit number i so that it may take out on the right-hand side of the reception device

specification 2 and may be shown.

[0022]Specification of the relay processing device in this transmission cue via which it goes, and the processing unit of a final destination, Supposing the reception device specification 1 - n are arranged from the transmission source processor in order even to the final destination processing unit, the BARIDDO bit v of a final destination processing unit and the relay processing device via which it goes shall be set to 1, it shall be effective, and the BARIDDO bit v of the other processing unit will be set to 0.

[0023]Drawing 6 is what showed other examples of the contents of the transmission cue of the main memory part 16 of drawing 3, and only the address processing unit specification 46 is formed in the transmission cue on main memory, and if it is in the example of drawing 6, as shown in drawing 5, the reception device specification 1 - n have. To this address processing unit specification 46, the reception device 48 by a logical address is stored. Corresponding to the reception device 48 which is a logical address stored in the address processing unit specification 46 of the transmission cue on main memory, the translator shown in drawing 7 is provided in the transmission treating part 18 of drawing 3.

[0024]If the conversion register 50 receives the reception device specification 48 from the transmission cue on the main memory shown in drawing 6 in drawing 7 as a logical address which shows the device number of a final destination device, With reference to the translation table 52 prepared beforehand, the real receiving set specification 0 - p are generated on the conversion register 50 as actual processing unit specification corresponding to the reception device specification 48 given with the logical address.

[0025]Since the device number of the final destination processing unit is known at this time, the BARIDDO bit v corresponding to the device number of a final destination processing unit and the relay processing device via which it goes until it results there is set to 1, respectively. Rewriting of each entry of the conversion register 50 shown in this drawing 7 comes to rewrite by a supervisor program from the instruction-processing part 14.

[0026]The main memory part of drawing 3 memorizes a command (program) and data, the instruction-processing part 14 takes out the command (program) from the main memory part 16, executive operation is carried out, and also the transmission treating part 18 is based on performing transmission processing of the packet between the main memory part 16 and the network 12. Next, the details of each part of the transmission treating part 18 shown in drawing 4 are explained.

[0027]In drawing 4, the control data transfer 26 is first started by the instructions from the instruction-processing part 14, With the value of the transmission cue read-out pointer 32 and the transmission cue writing pointer 34, start and transmission processing For read-out of the header unit of forward packets and body data, A main storage access demand is published to the main storage access control section 36, a network interface is controlled, and the data transfer from the data buffer 24 to the network 12 is controlled.

[0028]The main storage access control section 36 is accessed to the main memory part 16, and controls the data transfer between the main memory part 16 and the data buffer 24 by the instructions from the network transfer controlling part 26. When the data buffer 24 performs data transfer between a main memory part and the network 12, it buffers data temporarily, and. When writing the data on the main memory which the network transfer controlling part 26 and the main storage access control section 36 need, data is buffered temporarily.

[0029]The transmission cue writing pointer 34, the transmission cue read-out pointer 32, and the transmission cue base address 30 are registers in which the instruction controlling part 14 to reference and updating are possible. The transmission cue writing pointer 34 is for indicating whether the instruction-processing part 14 carried out the ENQ of the transfer request by the transmission cue throat top, and is for indicating whether the transmission cue read-out pointer 32 ended transmission processing by the transmission cue throat top.

[0030]Next, operation of the transmission treating part 18 is explained. First, the program executed in the instruction-processing part 14. It is the control information on the transfer request which includes the control information on reception device specification (address processing unit specification), body data length, a transmission address, a recipient address, and others in advance of the ENQ of a transfer request  $(\text{Transmission cue base address}) + (\text{transmission cue writing pointer}) \times (\text{header length})$

It comes out, and it is a data format of a packet header part, and as shown, for example in drawing 5, it writes in the position on the main memory shown. Then, the program of the instruction-processing part 14 \*\*\*\*\* the transmission cue writing pointer 34. Hereafter, the program of the instruction-processing part 14 is performed by repeating the writing of the control information on the transfer request to transmission cue, and increment of a transmission cue writing pointer, and ends ENQ processing.

[0031]The network transfer controlling part 26 is waiting for the ENQ of a transfer request in the state where it was started by the instructions from the instruction-processing part 14. If the transmission cue writing pointer 34 and a transmission cue read-out pointer stop being in agreement, it will be regarded as a thing with an unsettled transfer request, transmission processing will be started, and a main storage access demand will be published to the main storage access control section 36 for read-out of the header data of forward packets.

[0032]The main storage access control section 36 calculates the address on the main memory of the packet header data of the oldest unsettled transfer request from the value of the transmission cue base address register 30, and the value of the transmission cue read-out pointer 32, and publishes read-out access to the main memory part 16. If a packet header part is read from the main memory part 16, the main storage access control section 36 will store this in the data buffer 24, and it will notify the end of read-out of a packet header part to the network transfer controlling part 26.

[0033]The main storage access control section 36 calculates the address of packet body data by taking out the control information on the transfer request from the data buffer 24 succeeding, and publishes read-out access to the main memory part 16. If packet body data is read from the main memory part 16 one by one, the main storage access control section 36 will store this in the data buffer 24 one by one, and it will notify the amount of read-out of packet body data to the control data transfer 26 one by one.

[0034]If the end of read-out of a packet header part is notified to the control data transfer 26 from the main storage access control section 36, after it will take out the control information on a transfer request from the data buffer 24 and will add a necessary inspection and change, it returns control information to the data buffer 24. In the processing unit specified as a transmitting agency and a relay destination, a change of the packet header part by this control data transfer 26 will be made, if it is in this invention.

[0035]Then, the network transfer controlling part 26 sends out the packet header part with which an



inspection and change were able to be managed to the network 12. If the amount of read-out of packet body data is notified to the network transfer controlling part 26 one by one from the main storage access control section 36, it sends out the packet body data to have been stored in the data buffer 24 to the network 12 one by one.

[0036] In this way, after finishing sending out of the packet to the network 12, the network transfer controlling part 26 \*\*\*\*\* the transmission cue read-out pointer 32. Then, the network transfer controlling part 26 inspects the value of the transmission cue read-out pointer 32 and the transmission cue writing pointer 34, and if inharmonious, it will start the next transmission processing as that in which the unsettled transfer request remains. And this transmission processing is repeated until the value of the transmission cue read-out pointer 32 and the transmission cue writing pointer 34 becomes equal.

[0037] Drawing 8 is an explanatory view showing the original generation packet in the case of performing data transfer, without relaying other processing units from the processing unit 10-1 of the device number 1 shown in drawing 2 to the processing unit 10-2 of the device number 2 located in the next, and a actual transmitting packet. The left-hand side of drawing 8 is what showed the header unit 40 which exists in the transmission cue on main memory, and the body data 50 as one packet in all, i.e., original packet  $P_{00}$ , and all the reception device specification 1 - n are provided in the destination information field 42. The transmission source processor 1 is stored following this destination information field 42. Then, the course viewing area 54 is formed and the processing unit number via which it goes in the case of relay transfer is stored as course processing unit 2-n.

[0038] BARIDDO v shown in an initial stage by the leading bit of each specification of the destination information field 42 of original packet  $P_{00}$  is 0 altogether, and all of the course processing unit 2 - the BARIDDO bit of n have become 0 also about the course viewing area 54. In transmitting data to the processing unit 10-2 of the device number 2 as a final destination from the processing unit 10-1 of the device number 1, here, Transmitting packet  $P_{12}$  to the device number 2 moves the "processing unit specification 2" of the processing unit 10-2 of the device number 2 used as the following address to the head of the destination information field 42, as shown in right-hand side, and it sets the BARIDDO bit v to 1.

[0039] The "processing unit specification 1" which shows the transmission source processor which was in the head of the destination information field 42 by original packet  $P_{00}$  is moved to the last of the destination information field 42 as "the processing unit specification n+1." Although shifted on one about "the other processing unit specification 3 - the processing unit specification n", BARIDDO v is still 0. Packet  $P_{12}$  shown in the right-hand side of this drawing 8. If received by the processing unit 10-2 which is transmitted to the network 12 from the processing unit 10-1, and serves as an address. The transmission treating part 18 of the processing unit 10-2 investigates the destination information field 42 in the header unit 40 of packet  $P_{12}$  which received. When the BARIDDO bit v of the "processing unit specification 2" of a head is 1, the processing unit 10-2 is judged to be a packet made into an address, and according to directions of the header unit 40 after body data length, it comes to perform reception.

[0040] Drawing 9 goes via the device number 2 - (i-1) the processing unit 10-2 - 10-(i-1) in processing unit 10-i of the device number i from the processing unit 10-1 of the device number 1 of drawing 1. It is an explanatory view showing transmitting packet  $P_{12}$  to original packet  $P_{00}$  and the 2nd processing unit 10-2 in the transmission source processor 10-1 at the time of performing data transfer. Thus, when setting a final

destination to processing unit 10-i. At the same time it performs rearrangement in the destination information field 42 of transmitting packet  $P_{12}$  to the following processing unit 10-2 like the case of drawing 8. The BARIDDO bit v of the "processing unit specification i" of processing unit 10-i of the device number i used as a final destination is set to 1, and the BARIDDO bit v is altogether set to 1 also about "processing unit specification 2-(i-1)" of the relay processing device 10-2 - 10-(i-1) which are located between them.

[0041]Drawing 10 is an explanatory view showing receive-packet  $P_{12}$  in the 2nd processing unit 10-2 that received transmitting packet  $P_{12}$  shown in drawing 9 from the transmission source processor 10-1, and transmitting packet  $P_{23}$  to the 3rd following processing unit 10-3. If it is in the 2nd processing unit 10-2, the head of the destination information field 42 in the header data 40 of receive-packet  $P_{12}$  from a transmitting agency is referred to, From 1 standing on the BARIDDO bit v to the "processing unit specification 3" which 1 stands on the BARIDDO bit v of "the processing unit specification 2", and shows the following address. In this case, a self-processing unit recognizes it as it being not a final destination device but a device which performs relay, is rewritten like transmitting packet  $P_{23}$ , without carrying out reception of the receive-packet  $P_{12}$  with a self-processing unit, and is broadcast again to the following processing unit 10-3.

[0042]Rewriting of transmitting packet  $P_{23}$  in this case, The "processing unit specification 3" which shows the address of the 2nd next of receive-packet  $P_{12}$  about the destination information field 42 is moved to the head of transmitting packet  $P_{23}$ . The "processing unit specification 2" of the head of receive-packet  $P_{12}$  is moved to the last of the destination information field 42, and resets the BARIDDO bit v to 0. The other "processing unit specification 3" - "the processing unit specification n" is shifted on one.

[0043]It indicates that, and it put into transmitting packet  $P_{23}$ , and went via the processing unit 10-2 about the course viewing area 54 of receive-packet  $P_{12}$ . [ the BARIDDO bit v of the "course processing unit 2" in which the present processing unit 10-2 is shown ] Drawing 11 is an explanatory view of a receive packet and a transmitting packet in processing unit 10-(i-1) in front of [ of final destination device 10-i ] one.

[0044]In this case, even if it is, from the BARIDDO bit v of the head of the destination information field 42 in the header data 40 of receive-packet  $P_{(i-2)(i-1)}$  and the 2nd processing unit specification being set to 1. It changes and transmits to transmitting packet  $P_{(i-1)}$  which makes processing unit 10-i the following address without carrying out reception of a self-processing unit. As a result, in the destination information field 42 of transmitting packet  $P_{(i-1)(i)}$ , the BARIDDO bit is [ processing unit specification of a head ] 1. All other BARIDDO bits are 0.

[0045]It is adding to the last of the destination information field 42 as "processing unit specification n+i-1" which added n to the device number (i-1) of present processing unit 10-(i-1) in the last of the destination information field 42. About the course viewing area 54, it is shown that it is the device via which all the BARIDDO bits of "course processing unit 2" - "course processing unit i-1" already went by being set to 1. [ to present processing unit 10-(i-1) ]

[0046]If it is in processing unit 10-i used as a final destination, Receive packet  $P_{(i-1)(i)}$  shown in the right-hand side of drawing 11, and the BARIDDO bit v of the "processing unit specification i" of the head of the destination information field 42 in the header data 40 by 1. Since the BARIDDO bit v of the next "processing unit specification i+1" is 0, it judges that it is a packet which makes processing unit 10-i an address, and reception is performed according to the directions after the body data length following the course viewing area 54 in the header data 40.

[0047]Next, preferential control in case an intermediate processing unit relays is explained. If it is in the transmission treating part 18 when the packet which makes another processing unit a final destination is received, the transmission processing of the packet send data from the instruction-processing part 14 and relay data transfer processing in a self-processing unit may compete. Thus, when transmission and relay compete, it is > (relay processing) (transmitting processing), for example.

The data previously transmitted by the priority to say is determined. As another preferential control, the field which specifies the priority of data transfer is provided into the header unit of a packet, and it may be made to determine what looks at the priority of this appointed field and transmits previously.

[0048]For this reason, even if the relay data transfer to other processing units and send data transmission of a self-processing unit compete, according to predetermined turn, the output to the network 12 is appropriately controllable.

[0049]

[Effect of the Invention]If it is in the processing unit which performs each relay in the data transfer accompanied by the relay to the processing unit of the final destination which went via other processing units according to this invention as explained above, Since intervention of a supervisor program is not needed, reservation of the memory area on main memory is made unnecessary, the procedure of the data transfer accompanied by relay is simplified, and, simultaneously with reduction of a memory resource, high speed data transfer which reduced the overhead can be realized.

---

[Translation done.]

特開平6-301650

(43)公開日 平成6年(1994)10月28日

(51)Int.Cl. <sup>4</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 15/16	3 1 0 B	7429-5L		
13/00	3 5 5	7388-5B		
H 0 4 L 12/56		8529-5K	H 0 4 L 11/ 20	1 0 2 D
審査請求 未請求 請求項の数 5 O L (全 16 頁)				

(21)出願番号 特願平5-51186

(22)出願日 平成5年(1993)3月12日

(71)出願人 391037397

科学技術庁航空宇宙技術研究所長  
東京都調布市探大寺東町7丁目44番地1

(71)出願人 000005223

富士通株式会社  
神奈川県川崎市中原区上小田中1015番地

(72)発明者 三好 南

東京都八王子市横川町824-2

(72)発明者 福田 正大

東京都八王子市打越町1481-26

(74)代理人 弁理士 竹内 進 (外1名)

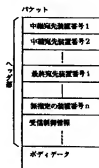
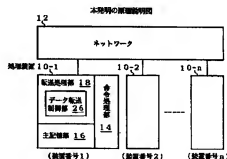
最終頁に続く

## (54)【発明の名称】 データ転送中継機構

## (57)【要約】

【目的】並列計算システムなどのデータ処理システムに設けた処理装置間で他の装置を経由して相互にデータを通信するデータ転送中継機構に関し、他の装置を経由する中継転送を効率的に行って通信手順の簡略化と処理のオーバーヘッドの低減を図る。

【構成】処理装置10-1～10-nの各々に、送信パケットのヘッダ部の中に最終宛先処理装置の装置番号に加えて1以上の経由処理装置の装置番号を指定し、受信パケットのヘッダデータの中の装置番号から自装置が最終宛先処理装置でなかった場合には、ボディデータを処理せずに次の経由処理装置を次の宛先とし、あるいは次の経由処理装置がなかった場合には最終宛先処理装置を次の宛先としてネットワーク12へパケット送り返すデータ転送制御部26を設ける。



## 【特許請求の範囲】

【請求項 1】 命令処理部（14）、主記憶部（16）

および転送処理部（18）を備えた複数の処理装置（10～1～10～n）をスイッチ回路を備えたネットワーク（12）に接続し、送信元処理装置の転送処理部（18）は、少なくとも宛先処理装置、転送データ賦、送信元処理装置におけるメモリアクセス属性、宛先処理装置における処理の指定に関する情報を含むヘッダ部と、該ヘッダ部で指定されるところのボディデータの量が零でない場合にはボディデータを少なくとも含むパケットを前記ネットワーク（12）に対して発信し、前記ネットワーク（12）のスイッチ回路は前記パケットに含まれる宛先処理装置の指示に従ってスイッチ回路の接続を行い、宛先処理装置の転送処理部（18）では、前記送信元処理装置が発信したボディデータの量が零でない場合にはヘッダ部に含まれる受信処理の制御情報を参照してボディデータを命令処理部（14）や主記憶部（16）に送って処理装置間の通信を行い、ボディデータの量が零の場合にはヘッダ部に含まれる制御情報を参照して処理装置間の通信を行うデータ処理システムに於いて、前記処理装置（10）の各々に、特定の処理装置に対する送信時には、ヘッダ部の中に最終宛先処理装置の装置番号に加えて 1 以上の経由処理装置の装置番号を指定したパケットをネットワーク（12）に送信し、ネットワーク（12）からのパケット受信時には該受信パケットのヘッダ部の中の装置番号から自装置が最終宛先処理装置でなかった場合には、ボディデータを処理せずに次の経由処理装置を次の宛先とし、あるいは次の経由処理装置がなかった場合には最終宛先処理装置を次の宛先としてネットワーク（12）へパケットを送り返す転送制御部（26）を設けたことを特徴とするデータ転送中継機構。

【請求項 2】 請求項 1 記載のデータ転送中継機構に於いて、送信元処理装置の転送制御部（22）は、ヘッダ部の中に最終宛先処理装置までの全ての経由処理装置の装置番号を格納したパケットを前記ネットワークに送信し、ヘッダ部の中に全ての経由処理装置の装置番号を含んだまま最終宛先処理装置にパケットを到達させることを特徴とするデータ転送中継機構。

【請求項 3】 請求項 1 又は 2 記載のデータ転送中継機構に於いて、前記転送制御部（22）は命令処理部（14）からのデータ転送処理要求があり、かつ同時に他の処理装置から自処理装置を経由するデータ転送を要求するようなパケットが送られてきているときには、自処理装置を中継するパケットの送信を優先することを特徴とするデータ転送中継機構。

【請求項 4】 請求項 1 又は 2 記載のデータ転送中継機構に於いて、前記パケットのヘッダデータの中にデータ転送の優先順位に関する情報を含み、前記転送制御部（22）は命令処理部（14）からのデータ転送処理要求と

同時に他の処理装置から自処理装置を経由するデータ転送を要求するようなパケットが送られてきているときには、転送の優先順位を比較し、中継パケットの優先順位が送信パケットの優先順位と同じか高いときに自処理装置を中継するパケットの送信を優先することを特徴とするデータ転送中継機構。

【請求項 5】 請求項 1 乃至 4 記載のデータ転送中継機構に於いて、前記転送制御部（22）は、主記憶部（16）上の転送キューにより指定された最終宛先処理装置の論理装置番号から最終宛先処理装置の実装置番号および経由処理装置の実装置番号を求める変換機構を備え、該変換機構により得られた最終宛先装置及び経由処理装置の実装置番号をヘッダ部の中に含むパケットを生成することを特徴とするデータ転送中継制御装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、並列計算機システムなどのデータ処理システムに設けた処理装置間で他の装置を経由して相互にデータを通信するデータ転送中継機構に関する。

【0002】

【従来の技術】 複数の処理装置をネットワークで接続したデータ処理システムにおける処理装置間のデータ通信において、2つの処理装置間のデータ転送経路に障害が発生した場合には、他の処理装置を経由してデータを転送する中継転送が行われる。例えば処理装置 A の記憶域の内容を他の処理装置 B に対して送出する場合、もし、処理装置 A から処理装置 B へ直接のデータ転送経路が、故障などの理由によって使用できないとき、処理装置 A は処理装置 C にデータ転送を行いその後処理装置 C から処理装置 B へデータ転送を行うことになる。

【0003】 またネットワークの特性によって、処理装置 A から処理装置 B へ直接のデータ転送経路が非常に複雑し、処理装置 A から処理装置 B の経路がすいている場合も、他の処理装置を経由する中継転送を行ったほうが効果的である。従来装置にあっては、処理装置 A から処理装置 C を経由して処理装置 B に送る中継転送について、処理装置 A のスーパーバイザプログラムが中継先の処理装置 C の主記憶に転送用のメモリ領域の確保を要求した後に、処理装置 A から処理装置 C にデータを転送して確保したメモリ領域にストアし、次に処理装置 C のスーパーバイザプログラムが処理装置 B の主記憶に転送用のメモリ領域の確保を要求した後に処理装置 B へのデータ転送を行っている。

【0004】

【発明が解決しようとする課題】 しかしながら、このような従来の中継転送では、送信元の処理装置 A のスーパーバイザプログラムが中継先の処理装置 C の主記憶にメモリ領域を確保してデータを転送し、処理装置 A から処理装置 C へのデータ転送の終了を待って処理装置 C から処

理装置Bへのデータ転送を行わなくてはならない。

【0005】このためデータ転送に要する処理のオーバヘッドが大きいし、メモリ資源も必要である。またメモリ資源を節約するためには、何らかに分けて転送しなくてはならないため、データ転送の手順が更に複雑になりオーバヘッドが非常に大きくなる。本発明は、このような従来の問題点に鑑みてなされたもので、他の装置を経由した中継転送を効率的に行って通信手順の簡略化と処理のオーバヘッドの低減を図るようにしたデータ転送中継機構に関する。

【0006】

【課題を解決するための手段】図1は本発明の原理説明図である。まず本発明は、命令処理部14、主記憶部16および転送処理部18を備えた複数の処理装置10-1〜10-nをスイッチ回路を備えたネットワーク12に接続したマルチプロセッサシステム等のデータ処理システムを対象とする。

【0007】送信元処理装置の転送処理部18は、少なくとも宛先処理装置、転送データ量、送信元処理装置におけるメモリアクセス属性、宛先処理装置における処理の指定に関する情報を含むヘッダ部と、ヘッダ部で指定されるところのボディデータとを少なくとも含むパケットをネットワーク12に対して送信する。ネットワーク12のスイッチ回路は送信パケットに含まれる宛先処理装置の指示に従ってスイッチ回路の接続を行う。

【0008】宛先処理装置の転送処理部18では送信元処理装置が発信したパケットのボディデータの量が零でない場合にヘッダ部に含まれる処理の情報を参照してボディデータを命令処理部14や主記憶部16に送って処理する。このような処理装置間の通信を行うデータ処理システムを対象とした本発明のデータ転送中継機構にあっては、処理装置10-1〜10-nの各々に、任意の処理装置への送信時に、ヘッダ部の中に最終宛先処理装置の装置番号に加えて1以上の経由処理装置の装置番号を指定したパケットをネットワーク12に送信し、ネットワーク12からのパケット受信時には、受信パケットのヘッダデータの中の実装置番号から自装置が最終宛先処理装置でなかった場合には、ボディデータを処理せず、次の経由処理装置を次の宛先とし、あるいは次の経由処理装置がなかった場合には最終宛先処理装置を次の宛先としてネットワーク12へパケットを送り返すデータ転送制御部26を設けたことを特徴とする。

【0009】ここで送信元処理装置のデータ転送制御部26は、ヘッダ部の中に最終宛先処理装置までの全ての経由処理装置の装置番号を格納したパケットをネットワーク12に送信し、ヘッダ部の中に全ての経由処理装置の装置番号を含んだまま最終宛先処理装置にパケットを到達させる。またデータ転送制御部26は命令処理部14からのデータ転送処理要求があり、かつ同時に他の処理装置から自処理装置を経由するデータ転送を要求する

ようなパケットが送られてきているときには、自処理装置を中継するパケットの送信を優先する。

【0010】更に、パケットのヘッダデータの中にデータ転送の優先順位に関する情報を含み、データ転送制御部26は命令処理部14からのデータ転送処理要求と同時に他の処理装置から自処理装置を経由するデータ転送を要求するようなパケットが送られてきているときには、転送の優先順位を比較し、中継パケットの優先順位が送信パケットの優先順位と同じか高いときに自処理装置を中継するパケットの送信を優先する。

【0011】一方、データ転送制御部26は、主記憶部16上の転送キュー20により指定された最終宛先処理装置の論理装置番号から最終宛先処理装置の実装置番号および経由処理装置の実装置番号を求めると交換機構を備え、この交換機構により得られた最終宛先装置及び経由処理装置の実装置番号をヘッダ部の中を含むパケットを生成するようにしてもよい。

【0012】

【作用】このような構成を備えた本発明のデータ転送中継機構によれば、次の作用が得られる。まず送信元処理装置のデータ転送制御部26にあっては、ヘッダ部(制御情報部)の中に最終宛先処理装置の装置番号のほかは経由する処理装置の装置番号を入れたパケットを生成してネットワーク12に発信する。

【0013】受信側の処理装置に設けたデータ転送制御部26は、受信パケットのヘッダ部がネットワーク12から受信データバッファ24に入ると、ヘッダ部の中の宛先指定番号を読んで自処理装置を中継して他処理装置を宛先とするパケットであるかどうか判別する。自処理装置を最終宛先とするパケットであるときには通常の受信処理を開始する。自処理装置を中継して他の処理装置を最終宛先とするパケットであるときには、ヘッダ部の中の現在の宛先処理装置の装置番号を次に受信する中継処理装置あるいは最終宛先処理装置の装置番号に直してから、ネットワークに再送信する。

【0014】また、データ転送制御部26で主記憶部16上の転送キューの論理装置番号から実処理装置番号への交換を行っているものでは、論理装置番号1つに対して中継する実処理装置番号および最終宛先の実処理装置番号を定義できる交換機構を設け、この交換機構を使用することで複雑な手順をもった中継データ転送をユーザプログラムで意識することなく高速に処理できる。

【0015】

【実施例】図2は本発明のデータ転送中継機構が適用されるマルチプロセッサシステムの実施例構成図であり、この実施例にあってはプロセッサエレメント(P/E)として知られた処理装置10-1、・・・10-nをネットワーク12に接続し、処理装置10-1〜10-nの相互間でネットワーク12を経由してデータ転送を行うようにしている。

【0016】即ち、処理装置10-1~10-nのそれぞれはヘッダ部とボディデータとで構成されたパケットを作成してネットワーク12に送信する。ここで、パケットのヘッダ部の中には少なくとも宛先処理装置の装置番号、転送データ量、送信元処理装置におけるメモリのアクセス属性、宛先処理装置における処理の指定に関する情報を含む。

【0017】ネットワーク12はクロスバースイッチや多段スイッチ回路網で構成され、処理装置10-1~10-nが発信したパケットに含まれる宛先処理装置の指示に従ってスイッチ接続を行う。ネットワーク12を経由してパケットを受信した処理装置は、パケットのヘッダ部に含まれる宛先処理装置における受信処理の制御情報を参照して受信処理を行う。

【0018】このような処理装置10-1~10-nの通常のパケット送受信に関する機能に加え、本発明においては、送信元処理装置にあっては送信パケットのヘッダ部の中に最終宛先処理装置の装置番号に加えて、1以上の経由処理装置の装置番号を入れ、また受信パケットのヘッダ部の中の装置番号から自処理装置が最終宛先処理装置でなかった場合にはボディデータを処理せずに次の経由処理装置を次の宛先とし、あるいは次の経由処理装置がなかった場合には最終宛先処理装置を次の宛先としてネットワークへパケットを送り返す機能が設けられる。

【0019】図3は図2の処理装置の実施例構成図であり、命令処理部14、主記憶部16及び転送処理部18を有する。主記憶部16にはパケット転送に用いる制御情報をパケットヘッダ部のデータ形式で格納した転送キュー20が設けられる。また転送処理部18には本発明のデータ中継転送機構としての機能を実現するデータ転送制御部26が設けられる。

【0020】図4は図3の転送処理部18の詳細を示した実施例構成図である。図4において、転送処理部18はデータバッファ24、データ転送制御部26、送信カウンタ28、転送キューベースアドレスレジスタ30、転送キュー読出ポインタ32、転送キュー書込ポインタ34及び主記憶アクセス制御部36で構成される。

【0021】図5は図3に示した主記憶部16上に作成される転送キュー20の内容を示す。図5において、主記憶上の転送キューにはパケット単位にヘッダ部のデータ形式でパケット転送制御に関する情報が格納され、図示の1パケット分のヘッダ部40を例にとると、宛先情報領域42に続いてボディデータ長、送信アドレス、受信アドレス等が設けられる。宛先情報領域42には図2に示した処理装置10-1~10-nの装置番号1~nが受信処理装置指定1~nとして登録されている。受信処理装置指定1~nのそれぞれは受信処理装置指定2の右側に取り出して示すように、バリッドビットvと処理装置番号iで構成されている。

【0022】この転送キューにおける経由する中継処理装置と最終宛先の処理装置の指定は、受信処理装置指定1~nが送信元処理装置から最終宛先処理装置まで順番に並べられているとすると、最終宛先処理装置及び経由する中継処理装置のバリッドビットを1にセットして有効とし、それ以外の処理装置のバリッドビットは0とする。

【0023】図6は図3の主記憶部16の転送キューの内容の他の例を示したもので、図6の実施例にあっては、主記憶上の転送キューには宛先処理装置指定46のみが設けられており、図5に示したように全ての受信処理装置指定1~nはもっていない。この宛先処理装置指定46に対しては、論理アドレスによる受信処理装置48が格納される。主記憶上の転送キューの宛先処理装置指定46に格納される論理アドレスである受信処理装置48に対応して、図3の転送処理部18には図7に示す交換機構が設けられている。

【0024】図7において、交換レジスタ50は図6に示した主記憶上の転送キューより最終宛先装置の装置番号を示す論理アドレスとして受信処理装置指定48を受けると、予め準備された交換テーブル52を参照し、論理アドレスで与えられた受信処理装置指定48に対応する実際の処理装置指定として受信装置指定0~pを交換レジスタ50上に生成する。

【0025】このとき、最終宛先処理装置の装置番号が分っていることから、最終宛先処理装置、及びそこに至るまでの経由する中継処理装置の装置番号に対応するバリッドビットvをそれぞれ1にセットする。この図7に示した交換レジスタ50の各エントリの書替えは、命令処理部14からスーパバイザプログラムにより書替えを行うようになる。

【0026】尚、図3の主記憶部は命令（プログラム）とデータを記憶するもので、また命令処理部14は主記憶部16からの命令（プログラム）を取り出して実行処理し、更に転送処理部18は主記憶部16とネットワーク12との間でパケットの転送処理を行うことを基本とする。次に図4に示した転送処理部18の各部の詳細を説明する。

【0027】図4において、まずデータ転送制御部26は命令処理部14からの指令によって起動され、転送キュー読出ポインタ32と転送キュー書込ポインタ34の値によって転送処理を開始し、転送パケットのヘッダ部とボディデータの読出しのため、主記憶アクセス制御部36へ主記憶アクセス要求を発行し、ネットワークインタフェースを制御し、データバッファ24からネットワーク12へのデータ転送を制御する。

【0028】主記憶アクセス制御部36はネットワーク転送制御部26からの指令によって主記憶部16に対しアクセスを行い、主記憶部16とデータバッファ24との間のデータ転送を制御する。データバッファ24は主

記憶部とネットワーク 12 との間でデータ転送を行う際に一時的にデータをバッファリングすると共に、ネットワーク転送制御部 26 と主記憶アクセス制御部 36 が必要とする主記憶上のデータの読み書きを行う際に一時的にデータをバッファリングするものである。

【0029】転送キュー書込ポインタ 34、転送キュー読出ポインタ 32 及び転送キューベースアドレス 30 は命令制御部 14 から参照及び更新可能なレジスタであって、転送キュー書込ポインタ 34 は命令処理部 14 が転送キューのどこまで転送要求をエンキューしたかを表示するためのものであり、転送キュー読出ポインタ 32 は転送キューのどこまで転送処理を終了したかを表示するためのものである。

【0030】次に転送処理部 18 の動作を説明する。まず、命令処理部 14 で実行するプログラムは、転送要求のエンキューに先立って受信処理装置指定（宛先処理装置指定）、ボディータ長、送信アドレス、受信アドレス及びその他の制御情報を含む転送要求の制御情報を、（転送キューベースアドレス）+（転送キュー書込ポインタ）×（ヘッダ長）

で示される主記憶上の位置にパケットヘッダ部のデータ形式で、例えば図 5 に示すように書き込む。続いて命令処理部 14 のプログラムは、転送キュー書込ポインタ 34 をインクリメントする。以下、命令処理部 14 のプログラムは転送キューへの転送要求の制御情報の書込みと、転送キュー書込ポインタのインクリメントとを繰り返して行い、エンキュー処理を終了する。

【0031】ネットワーク転送制御部 26 は命令処理部 14 からの指令によって起動された状態で転送要求のエンキューを待っており、転送キュー書込ポインタ 34 と転送キュー読出ポインタとが一致しなくなると未処理の転送要求があるものと見做して転送処理を開始し、転送パケットのヘッダデータの読出のため、主記憶アクセス制御部 36へ主記憶アクセス要求を発行する。

【0032】主記憶アクセス制御部 36 は転送キューベースアドレスレジスタ 30 の値で転送キュー読出ポインタ 32 の値から最も古い未処理の転送要求のパケットヘッダデータの主記憶上のアドレスを計算し、主記憶部 16 に対し読出アクセスを発行する。主記憶部 16 からパケットヘッダ部が読み出されてくると、主記憶アクセス制御部 36 はこれをデータバッファ 24 に格納すると共に、ネットワーク転送制御部 26 に対しパケットヘッダ部の読出終了を通知する。

【0033】主記憶アクセス制御部 36 は、引き続いてデータバッファ 24 からの転送要求の制御情報を取り出してパケットボディータのアドレスを計算し、主記憶部 16 に対し読出アクセスを発行する。主記憶部 16 からパケットボディータが順次読み出されてくると、主記憶アクセス制御部 36 はこれを順次データバッファ 24 に格納すると共に、データ転送制御部 26 にパケット

ボディータの読出量を順次通知する。

【0034】データ転送制御部 26 は主記憶アクセス制御部 36 からパケットヘッダ部の読出終了を通知されると、データバッファ 24 から転送要求の制御情報を取り出して所要の検査と変更を加えてから、制御情報をデータバッファ 24 に書き直す。このデータ転送制御部 26 によるパケットヘッダ部の変更は、本発明にあっては、送信元及び中継先として指定された処理装置において行われることになる。

【0035】引き続いて、ネットワーク転送制御部 26 は検査及び変更が済んだパケットヘッダ部をネットワーク 12 に送出する。更にネットワーク転送制御部 26 は主記憶アクセス制御部 36 からパケットボディータの読出量を順次通知されると、データバッファ 24 に格納された分のパケットボディータを順次ネットワーク 12へ送出する。

【0036】こうしてネットワーク 12 へのパケットの送出を終えると、ネットワーク転送制御部 26 は転送キュー読出ポインタ 32 をインクリメントする。引き続いて、ネットワーク転送制御部 26 は転送キュー読出ポインタ 32 と転送キュー書込ポインタ 34 の値を検査し、不一致であれば未処理の転送要求が残っているものとして次の転送処理を開始する。そして、この転送処理を転送キュー読出ポインタ 32 と転送キュー書込ポインタ 34 の値が等しくなるまで繰り返す。

【0037】図 8 は図 2 に示した装置番号 1 の処理装置 10-1 から次に位置する装置番号 2 の処理装置 10-2 に対し、他の処理装置を中継せずにデータ転送を行う場合のオリジナルの生成パケットと実際の送信パケットを示した説明図である。図 8 の左側は主記憶上の転送キューに存在するヘッダ部 40 とボディータ 50 を合わせて 1 つのパケット即ちオリジナルのパケット P<sub>00</sub> として示したもので、宛先情報領域 42 には全ての受信処理装置指定 1〜n が設けられる。この宛先情報領域 42 に続いて送信元処理装置 1 が格納される。続いて経由情報領域 54 が設けられ、中継転送の際に経由する処理装置番号が経由処理装置 2〜n として格納されている。

【0038】初期段階でオリジナルパケット P<sub>00</sub> の宛先情報領域 42 の各指定の先頭ビットで示すバリッド v は全て 0 となっており、また経由情報領域 54 については、經由処理装置 2〜n のバリッドビットは全て 0 となっている。ここで、装置番号 1 の処理装置 10-1 から最終宛先として装置番号 2 の処理装置 10-2 にデータを転送する場合には、装置番号 2 への送信パケット P<sub>12</sub> は右側に示すように次の宛先となる装置番号 2 の処理装置 10-2 の「処理装置指定 2」を宛先情報領域 42 の先頭に移し、バリッドビット v を 1 に設定する。

【0039】また、オリジナルパケット P<sub>00</sub> で宛先情報領域 42 の先頭にあった送信元処理装置を示す「処理装置指定 1」は、宛先情報領域 42 の最後に「処理装置指



定 $n+1$ として移す。それ以外の「処理装置指定3～処理装置指定 $n$ 」については1つ上にずらす、バリッド $v$ は0のままである。この図8の右側に示すパケット $P_{12}$ は、処理装置10-1からネットワーク12に送信され、宛先となる処理装置10-2で受信されると、処理装置10-2の転送処理部18は受信したパケット $P_{12}$ のヘッダ部40の中の宛先情報領域42を調べ、先頭の「処理装置指定2」のバリッドビット $v$ のみが1であるときには処理装置10-2を宛先とするパケットであると判断し、ボディデータ長以降のヘッダ部40の指示に従って受信処理を行うようになる。

【0040】図9は図1の装置番号1の処理装置10-1から装置番号1の処理装置10-1に装置番号2～(1-1)の処理装置10-2～10-(1-1)を経由してデータ転送を行う際の送信元処理装置10-1におけるオリジナルパケット $P_{00}$ と2番目の処理装置10-2への送信パケット $P_{12}$ を示した説明図である。このように、最終宛先は処理装置10-1とするときには、図8の場合と同様にして次の処理装置10-2への送信パケット $P_{12}$ の宛先情報領域42における並び替えを行うと同時に、最終宛先となる装置番号1の処理装置10-1の「処理装置指定1」のバリッドビット $v$ を1にセットすると共に、その間に位置する中継処理装置10-2～10-(1-1)の「処理装置指定2～(1-1)」についても全てバリッドビット $v$ を1にセットする。

【0041】図10は図9に示す送信パケット $P_{12}$ を送信元処理装置10-1より受信した2番目の処理装置10-2における受信パケット $P_{12}$ と次の3番目の処理装置10-3に対する送信パケット $P_{23}$ を示した説明図である。2番目の処理装置10-2にあっては、送信元からの受信パケット $P_{12}$ のヘッダデータ40の中の宛先情報領域42の先頭を参照し、「処理装置指定2」のバリッドビット $v$ に1が立っており、且つ次の宛先を示す「処理装置指定3」にもバリッドビット $v$ に1が立っていることから、この場合には自処理装置は最終宛先装置ではなく、中継を行う装置であると認識し、受信パケット $P_{12}$ を自処理装置で受信処理せず、送信パケット $P_{23}$ のように書き替えて、次の処理装置10-3に再送信する。

【0042】この場合の送信パケット $P_{23}$ の書替は、宛先情報領域42については受信パケット $P_{12}$ の2番目の次の宛先を示す「処理装置指定3」を送信パケット $P_{23}$ の先頭に移し、受信パケット $P_{12}$ の先頭の「処理装置指定2」は宛先情報領域42の最後に移してバリッドビット $v$ を0にリセットする。それ以外の「処理装置指定3」～「処理装置指定 $n$ 」は1つ上にずらす。

【0043】更に、受信パケット $P_{12}$ の經由表示領域54について、現在の処理装置10-2を示す「經由処理装置2」のバリッドビット $v$ をオンして送信パケット $P_{23}$

23に入れ、処理装置10-2を経由したことを表示する。図11は最終宛先装置10-1の1つ前の処理装置10-(1-1)における受信パケットと送信パケットの説明図である。

【0044】この場合にあっても、受信パケット $P_{(i-1)-(i-1)}$ のヘッダデータ40の中の宛先情報領域42の先頭と2番目の処理装置指定のバリッドビット $v$ が1にセットされていることから、自処理装置の受信処理をせずに処理装置10-1を次の宛先とする送信パケット $P_{(i-1)-1}$ に変更して送信する。この結果、送信パケット $P_{(i-1)-(i-1)}$ の宛先情報領域42は、先頭の処理装置指定のみがバリッドビットが1となっており、他のバリッドビットは全て0となっている。

【0045】また、宛先情報領域42の最後には、現在の処理装置10-(1-1)の装置番号(1-1)に $n$ を加えた「処理装置指定 $n+1-1$ 」として宛先情報領域42の最後に加えている。更に、經由表示領域54については、現在の処理装置10-(1-1)までの「經由処理装置2」～「經由処理装置1-1」のバリッドビットが全て1となっており、既に經由した装置であることを示している。

【0046】最終宛先となる処理装置10-1にあっては、図11の右側に示したパケット $P_{(i-1)-1}$ を受信し、ヘッダデータ40の中の宛先情報領域42の先頭の「処理装置指定1」のバリッドビット $v$ が1で、次の「処理装置指定1+1」のバリッドビット $v$ が0であることから、処理装置10-1を宛先とするパケットであると判断し、ヘッダデータ40の中の經由表示領域54に続くボディデータ長以降の指示に従って受信処理を行う。

【0047】次に途中の処理装置で中継を行う場合の優先制御を説明する。別の処理装置を最終宛先とするパケットを受信した際に、転送処理部18にあっては、自処理装置内の命令処理部14からのパケット送信データの転送処理と中継データの転送処理が競合する場合がある。このように送信と中継が競合した場合には、例えば、

(中継処理) > (送信処理)

という優先順位で先に送信するデータを決定する。また、別な優先制御としては、パケットのヘッダ部の中にデータ転送の優先順位を指定するフィールドを設け、この指定フィールドの優先順位を見て先に送信するものを決定するようにしてもよい。

【0048】このため、他の処理装置への中継データ転送と自処理装置の送信データ転送とが競合しても、所定の順番に従って適切にネットワーク12に対する出力を制御することができる。

【0049】

【発明の効果】以上説明してきたように本発明によれば、他の処理装置を経由した最終宛先の処理装置への中

継を伴うデータ転送において、各中継を行う処理装置にあっては、スーパーバイザプログラムの介入を必要としないことから主記憶上のメモリ領域の確保を不要とし、メモリ資源の節減と同時に、中継を伴うデータ転送の手順を簡単にし、オーバーヘッドを低減した高速データ転送を実現できる。

【図面の簡単な説明】

【図1】本発明の原理説明図

【図2】本発明が適用されるマルチプロセッサシステムの実施例構成図

【図3】図1の処理装置の実施例構成図

【図4】図3の転送処理部の詳細を示した実施例構成図

【図5】主記憶上の転送キューの説明図

【図6】主記憶上の転送キューの他の例を示した説明図

【図7】図6の主記憶上の転送キューで指定した論理装置番号を実装置番号に変換する変換機構の説明図

【図8】装置番号1から装置番号2に直接転送する場合のバケット説明図

【図9】装置番号1から装置番号1に中継転送する場合の送信元バケットの説明図

【図10】装置番号2の処理装置における中継時の受信、送信バケットの説明図

【図11】最終宛先装置番号1のより1つ前の処理装置における中継時の受信、送信バケットの説明図

【符号の説明】

10-1~10-n: 処理装置

12: ネットワーク

14: 命令処理部

16: 主記憶部

18: 転送処理部

20: 転送キュー

24: データバッファ

26: データ転送制御部

28: 送信カウンタ

30: 転送キューベースアドレスレジスタ

32: 転送キュー読出ポインタ

34: 転送キュー書込ポインタ

36: 主記憶アクセス制御部

40: 1バケット分のヘッダ部

42: 宛先情報領域

48: 論理受信処理装置指定

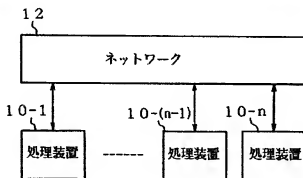
50: 変換レジスタ

50: テーブル

60: ボディデータ

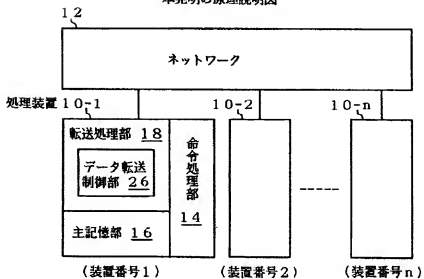
【図2】

本発明が適用されるマルチプロセッサシステムの実施例構成図

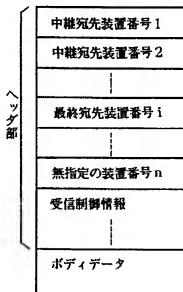


【図1】

## 本発明の原理説明図

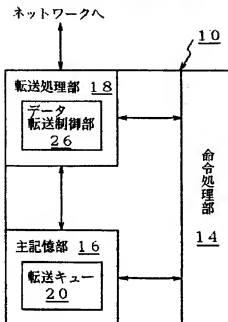


## パケット



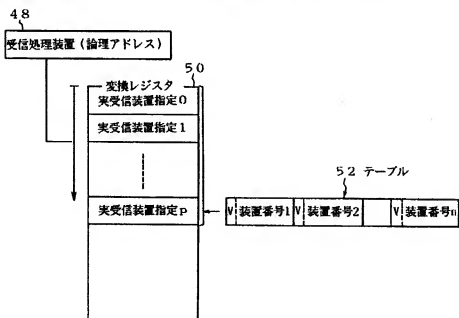
【図3】

図1の処理装置の実施例構成図



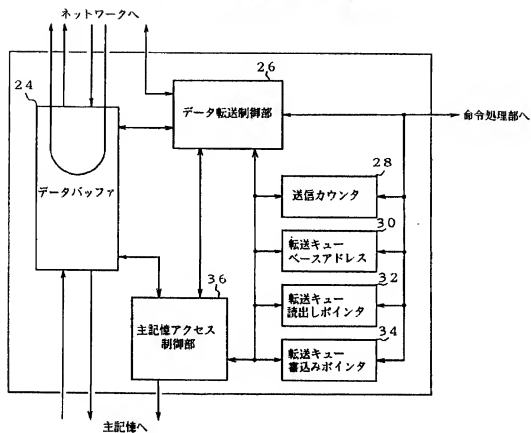
【図7】

図6の主記憶上の転送キューで指定した論理装置番号を実装置番号に変換する変換機構の説明図



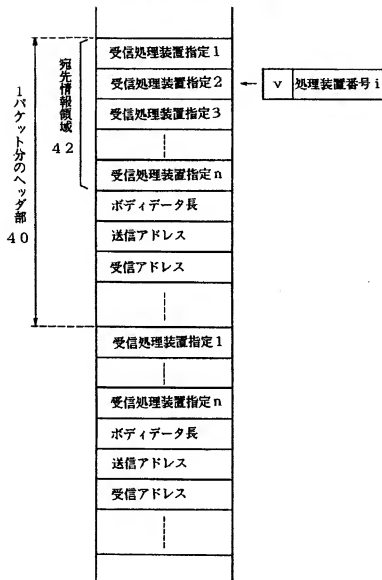
【図4】

図3の転送処理部の詳細を示した実施例構成図



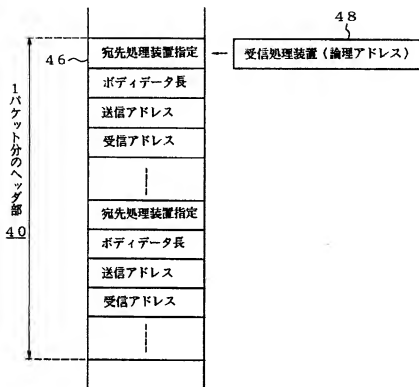
【図5】

主記憶上の転送キューの説明図



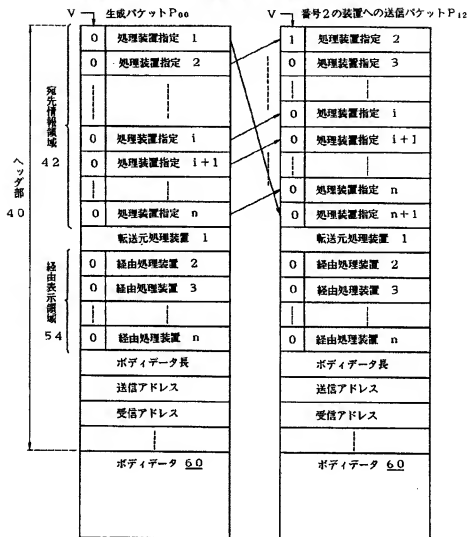
【図 6】

主記憶上の転送キューの他の例を示した説明図



【図8】

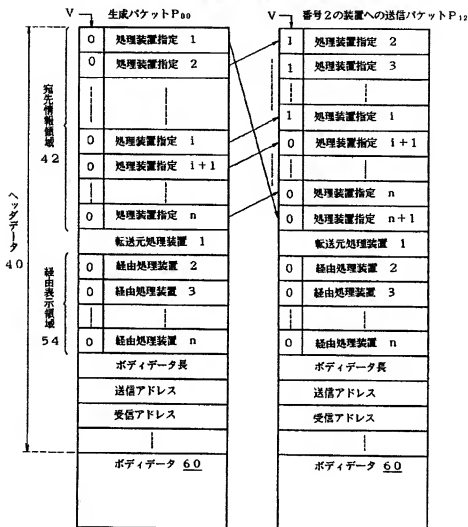
装置番号1から装置番号2に直接転送する場合のパケット説明図





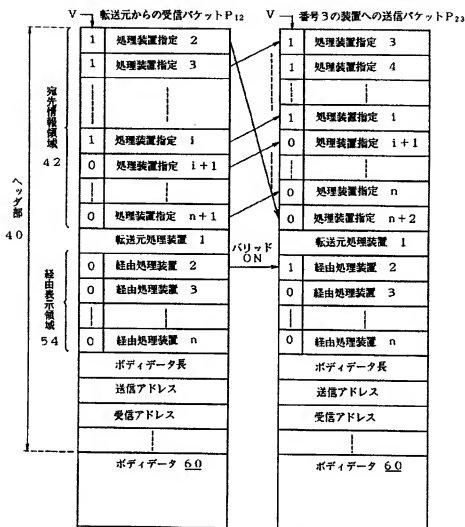
【図9】

装置番号1から装置番号iに中継転送する場合の送信元パケットの説明図



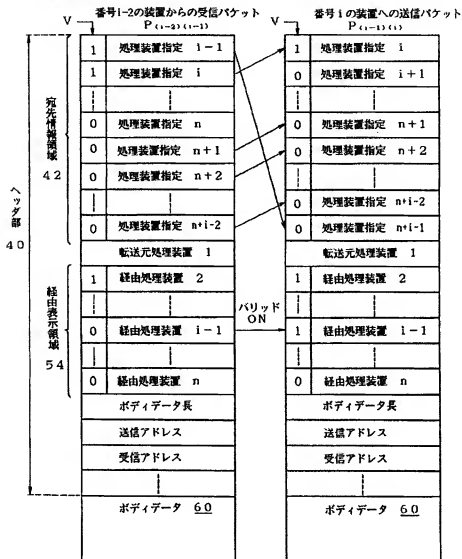
【図10】

装置番号2の処理装置における中継時の受信、送信パケットの説明図



【図11】

最終宛先装置番号1のより1つ前の処理装置における中継時の受信、送信パケットの説明図



フロントページの続き

(72)発明者 池田 正幸  
神奈川県川崎市中原区小田中1015番地  
富士通株式会社内

(72)発明者 高村 守幸  
神奈川県川崎市中原区小田中1015番地  
富士通株式会社内